Genotype Imputation and Variability in Polygenic Risk Score Estimation

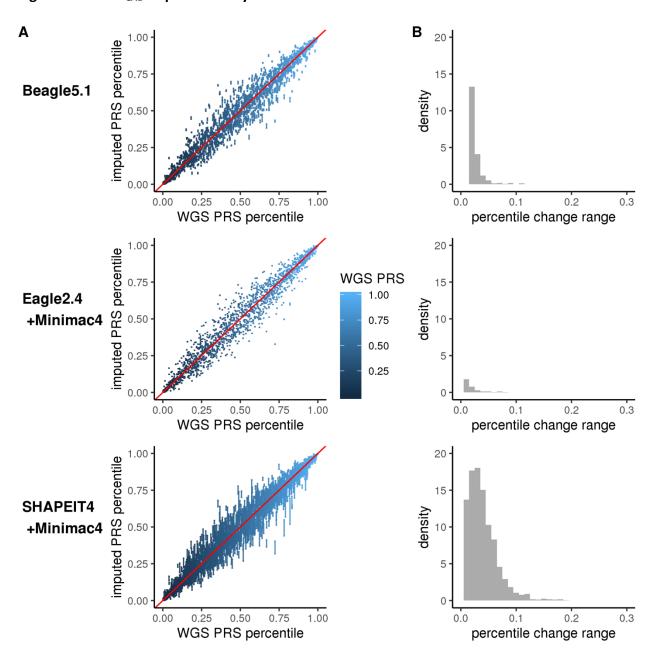
Shang-Fu Chen^{1,2}, Raquel Dias^{1,2}, Doug Evans^{1,2}, Elias L. Salfati^{1,2}, Shuchen Liu^{1,2}, Nathan E Wineinger^{1,2}, Ali Torkamani^{1,2}

Corresponding Author: atorkama@scripps.edu

¹ Scripps Research Translational Institute, La Jolla, California, 92037, USA

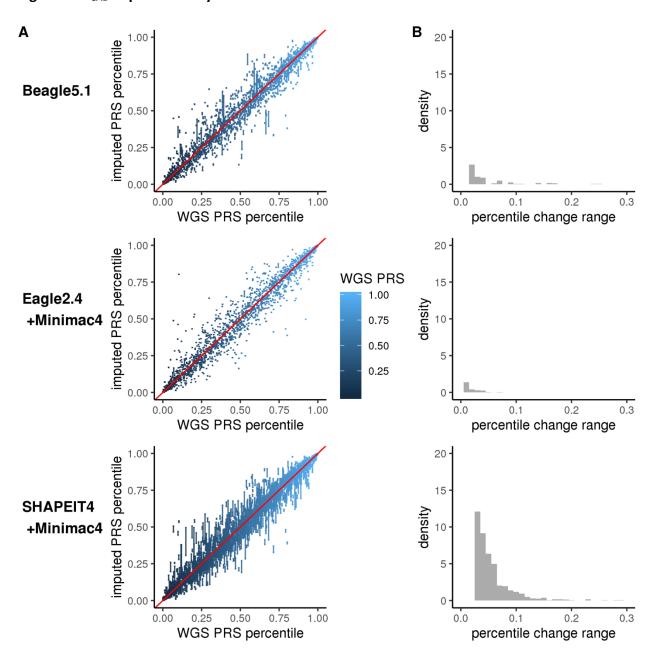
 $^{^2} Department \ of \ Integrative \ Structural \ and \ Computational \ Biology, Scripps \ Research, La \ Jolla, California, 92037, USA$

Fig S1. metaGRS_{CAD} Reproducibility.



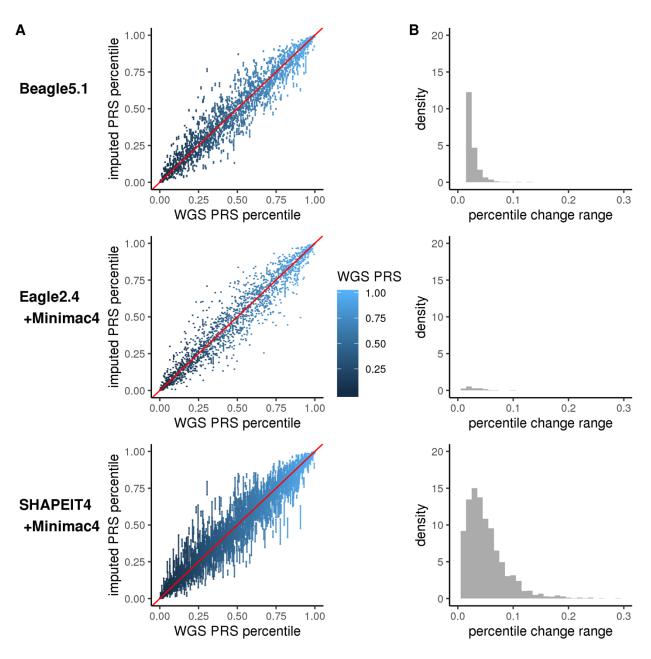
The variability in metaGRS_{CAD} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S2. GPS_{CAD} Reproducibility.



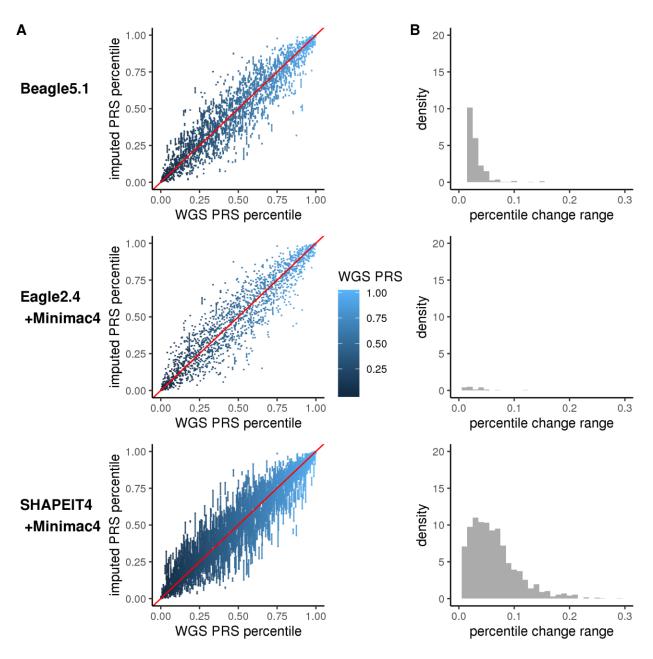
The variability in GPS_{CAD} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S3. PRS-GWAS_{T2D} (547) Reproducibility.



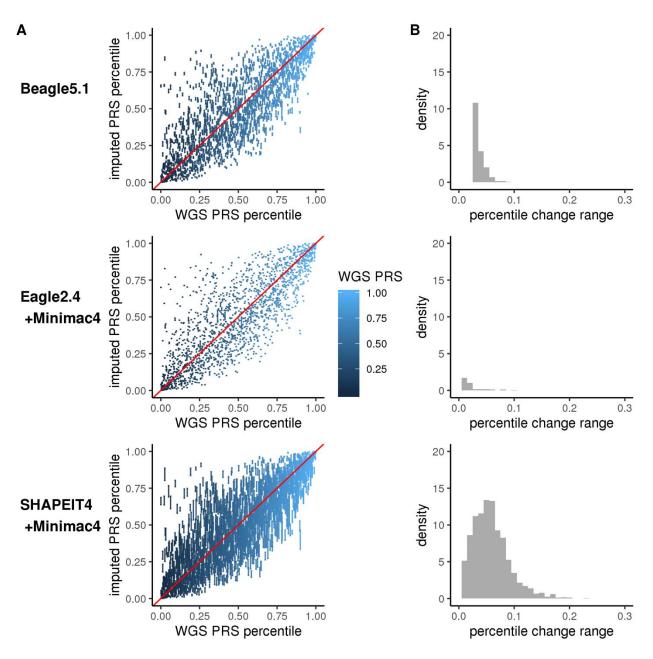
The variability in PRS-GWAS_{T2D} (547) percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S4. PRS-GWAS_{T2D} (397) Reproducibility.



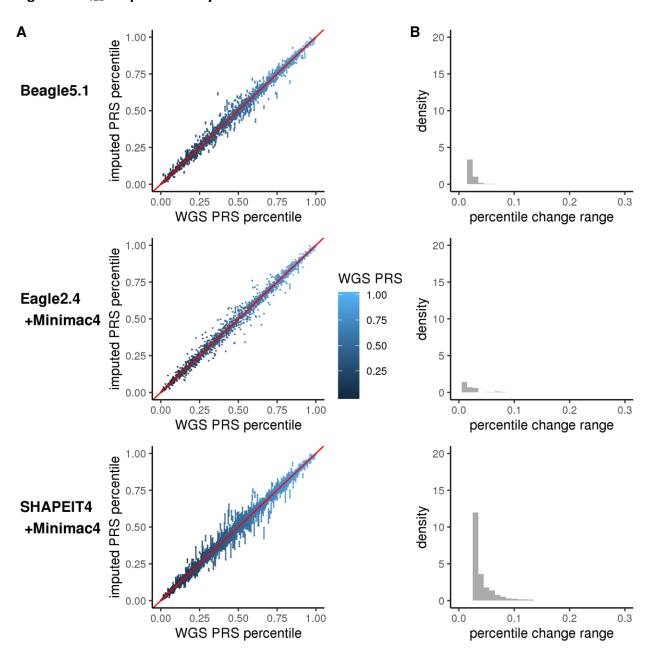
The variability in PRS-GWAS_{T2D} (397) percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S5. PRS-GWAS_{T2D} (170487) Reproducibility.



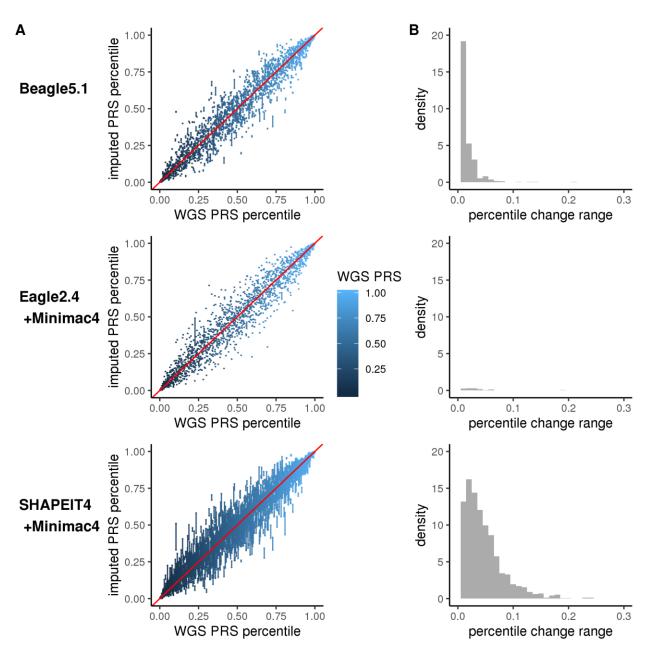
The variability in PRS-GWAS_{T2D} (170487) percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S6. GPS_{T2D} Reproducibility.



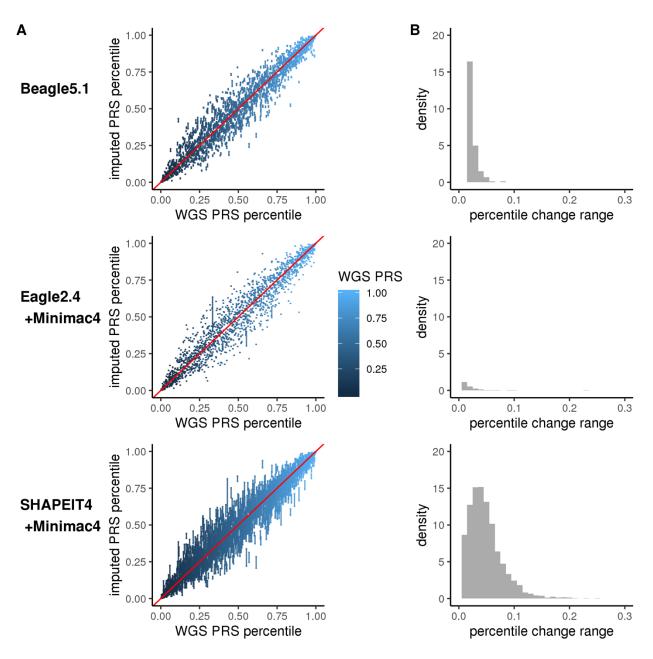
The variability in GPS_{T2D} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S7. PRS-GWAS_{BC} (239) Reproducibility.



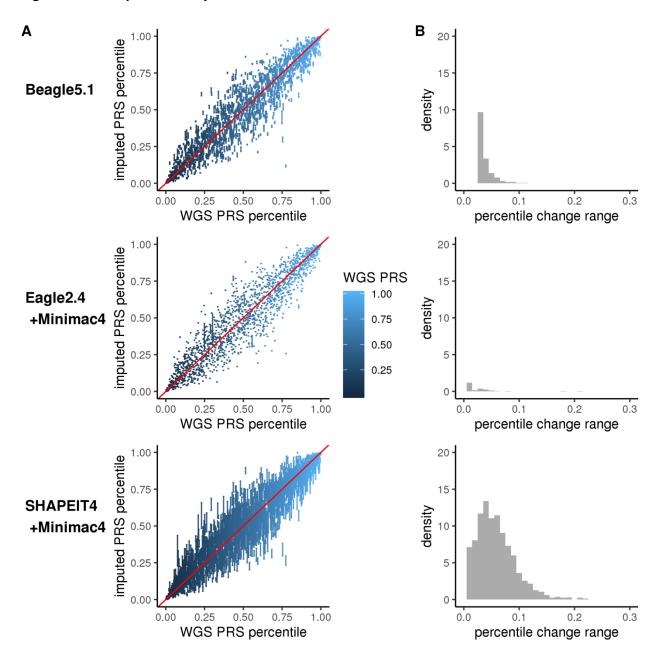
The variability in PRS-GWAS_{BC} (239) percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S8. PRS-GWAS_{BC} (2935) Reproducibility.



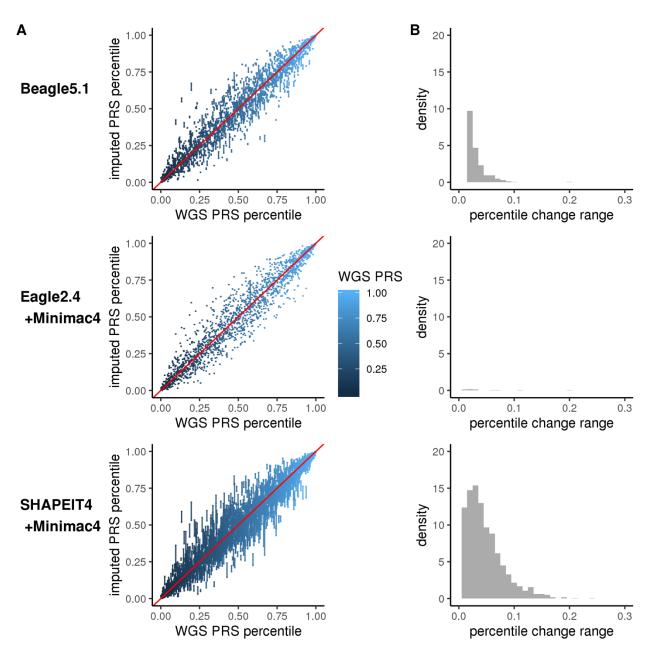
The variability in PRS-GWAS_{BC} (2935) percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S9. GPS_{BC} Reproducibility.



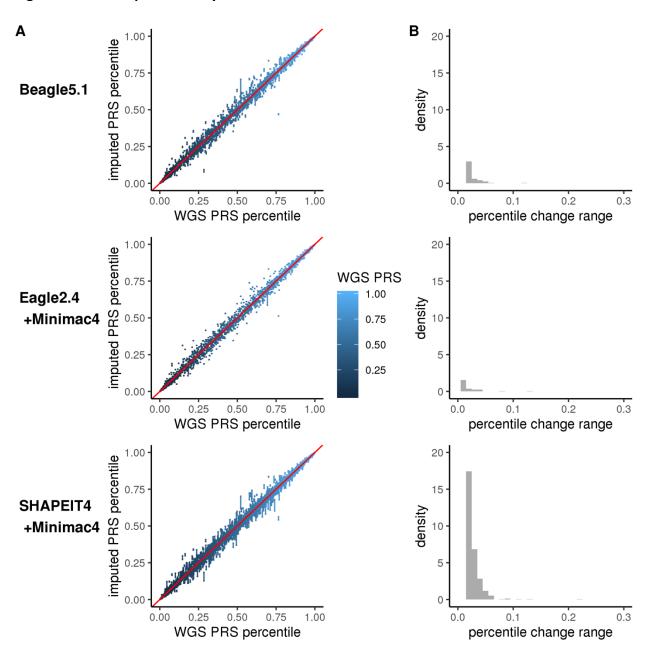
The variability in GPS_{BC} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S10. PRS-GWAS_{Afib} Reproducibility.



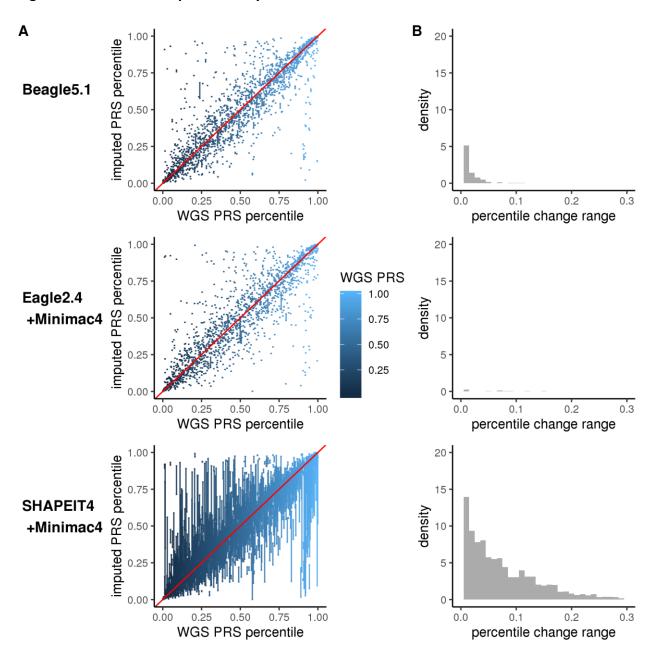
The variability in PRS-GWAS_{Afib} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S11. GPS_{Afib} Reproducibility.



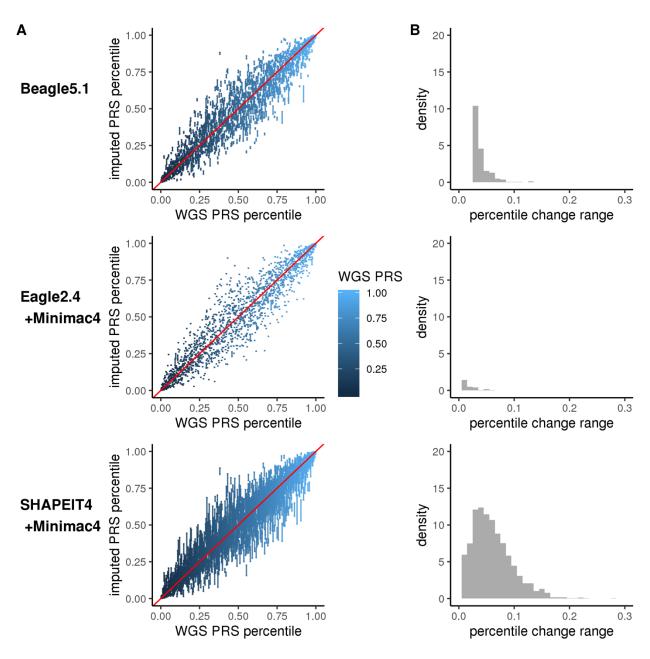
The variability in GPS_{Afib} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S12. PRS-GWAS_{AD} Reproducibility.



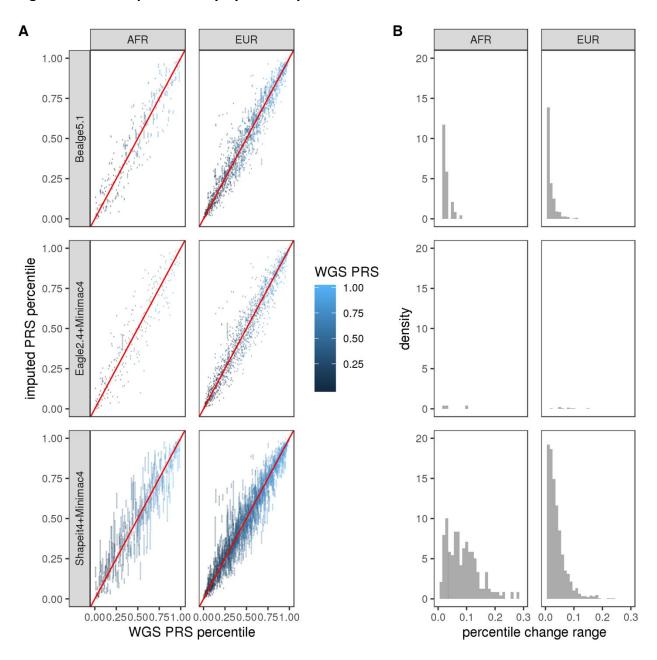
The variability in PRS-GWAS_{AD} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S13. PRS-GWAS_{Glaucoma} Reproducibility.



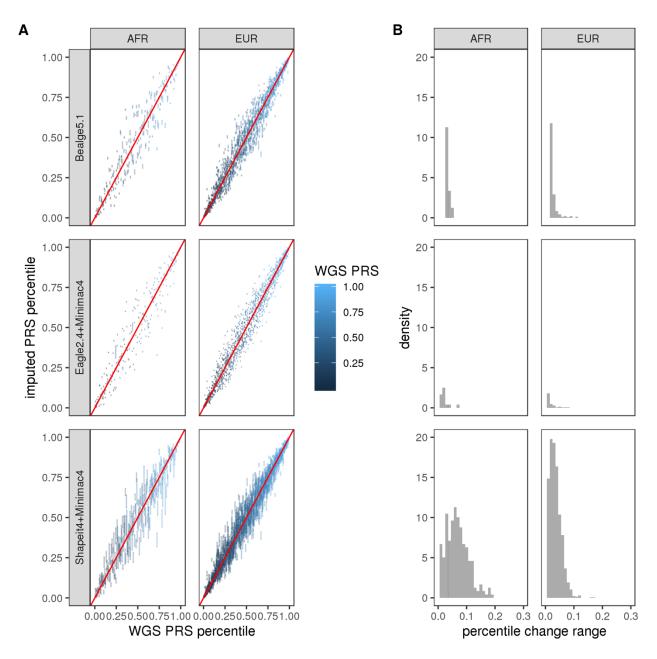
The variability in PRS-GWAS_{Glaucoma} percentile values as determined by three different imputation processes. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown.

Fig S14. PRS_{CAD} Reproducibility by Ancestry.



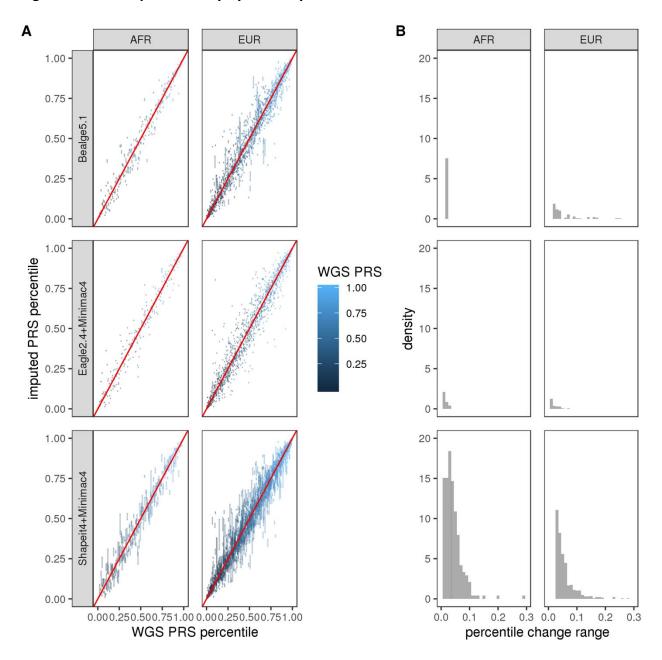
The variability in PRS_{CAD} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S15. metaGRS_{CAD} Reproducibility by Ancestry.



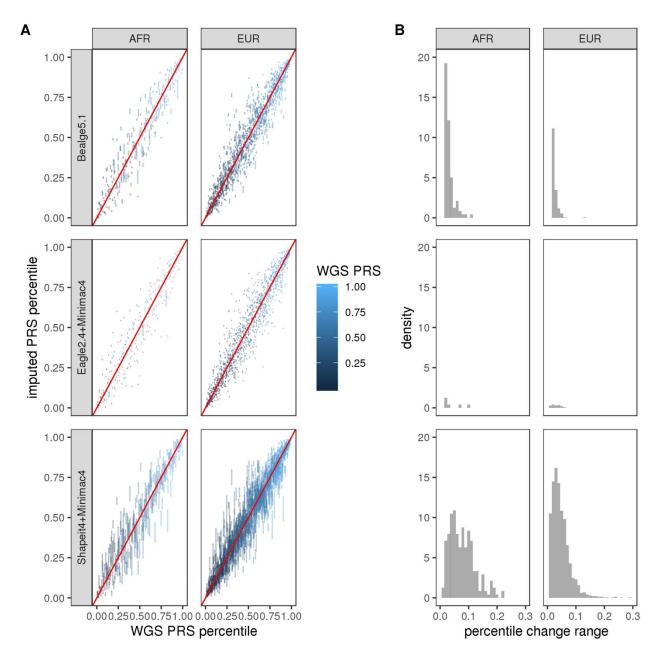
The variability in metaGRS_{CAD} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S16. GPS_{CAD} Reproducibility by Ancestry.



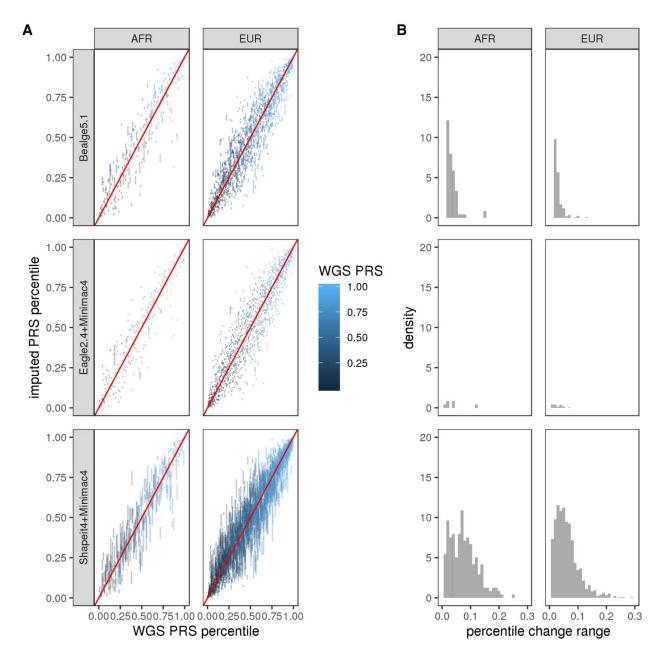
The variability in GPS_{CAD} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S17. PRS-GWAS_{T2D} (547) Reproducibility by Ancestry.



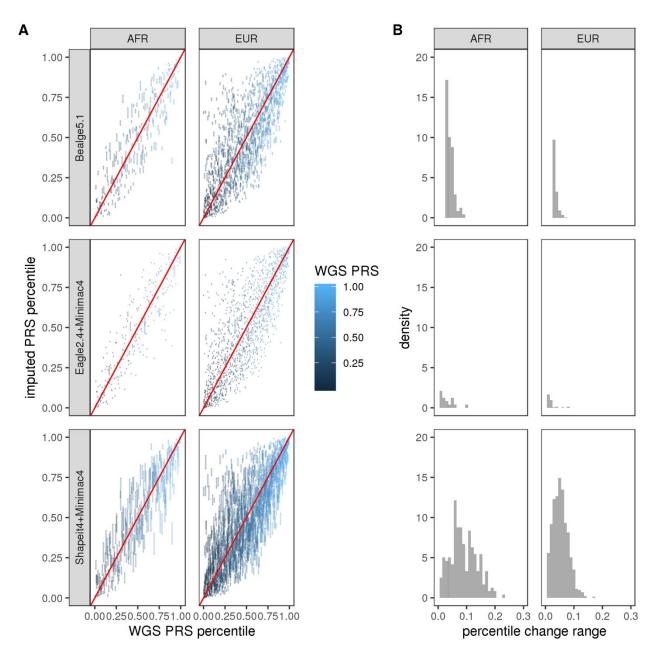
The variability in PRS-GWAS_{T2D} (547) percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S18. PRS-GWAS_{T2D} (397) Reproducibility by Ancestry.



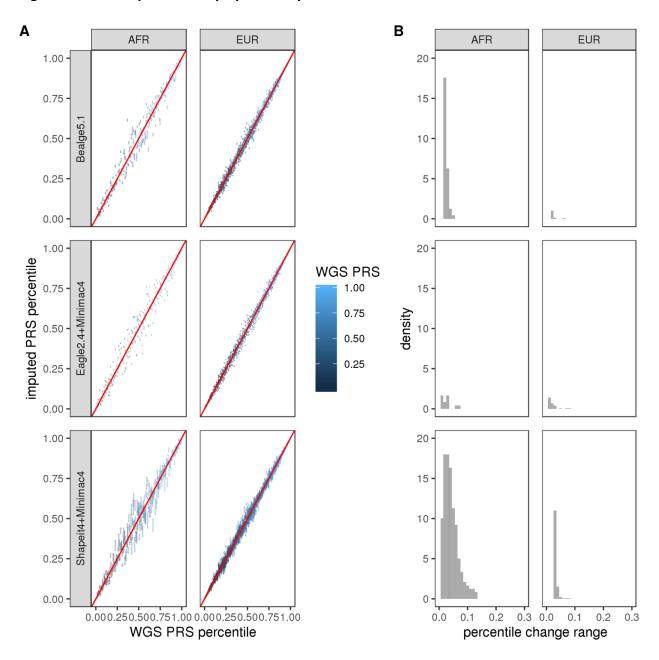
The variability in PRS-GWAS_{T2D} (397) percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S19. PRS-GWAS_{T2D} (170487) Reproducibility by Ancestry.



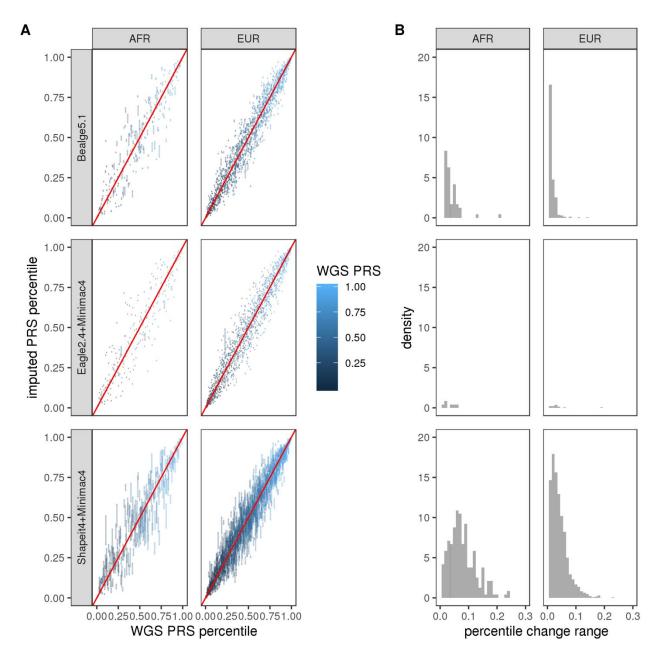
The variability in PRS-GWAS_{T2D} (170487) percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S20. GPS_{T2D} Reproducibility by Ancestry.



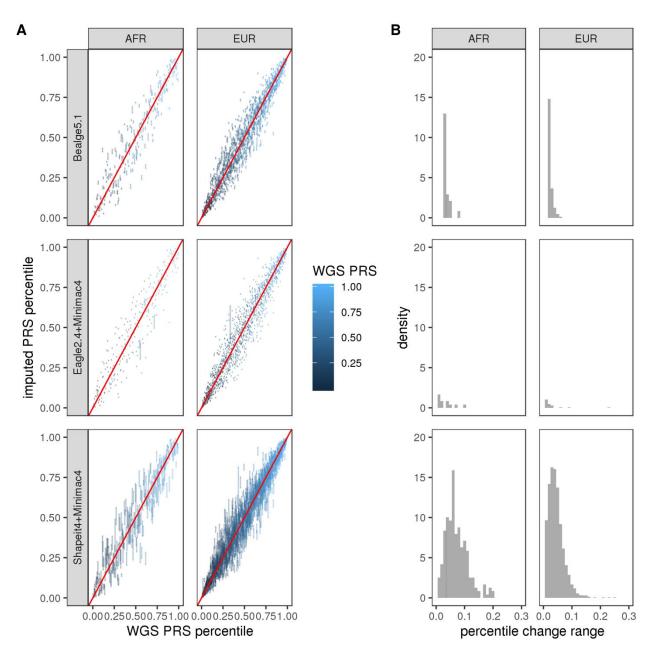
The variability in GPS_{T2D} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S21. PRS-GWAS_{BC} (239) Reproducibility by Ancestry.



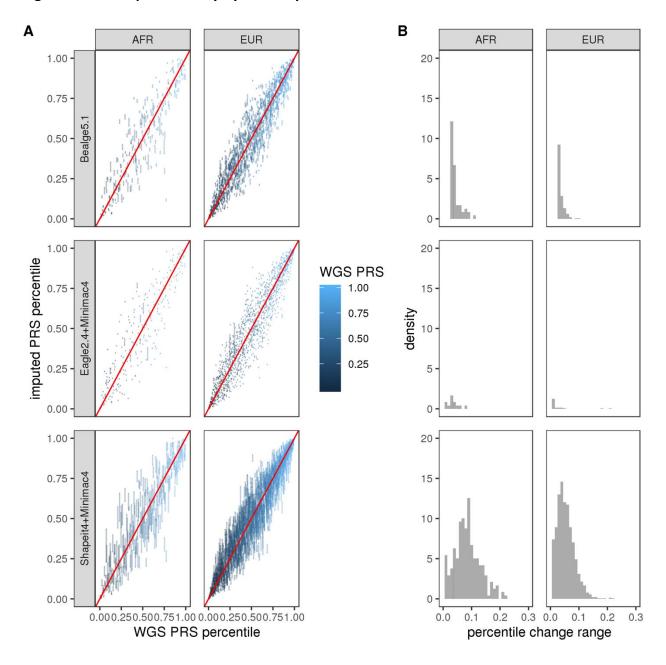
The variability in PRS-GWAS_{BC} (239) percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S22. PRS-GWAS_{BC} (2935) Reproducibility by Ancestry.



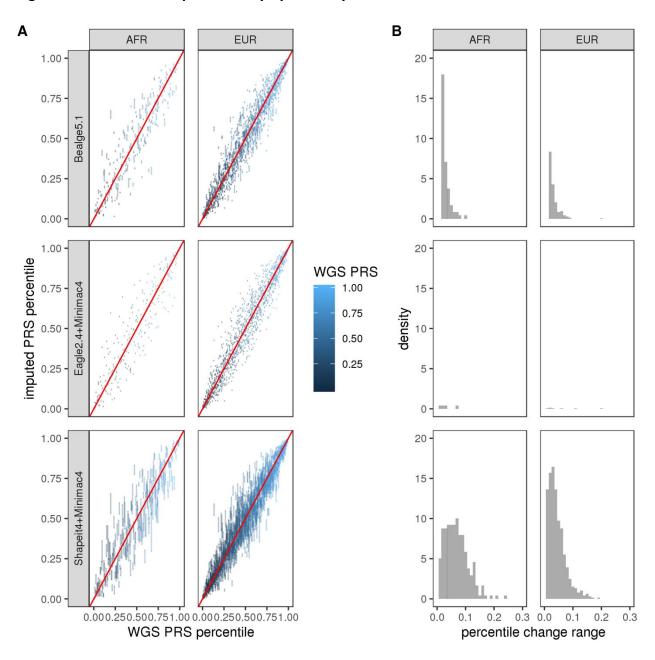
The variability in PRS-GWAS_{BC} (2935) percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S23. GPS_{BC} Reproducibility by Ancestry.



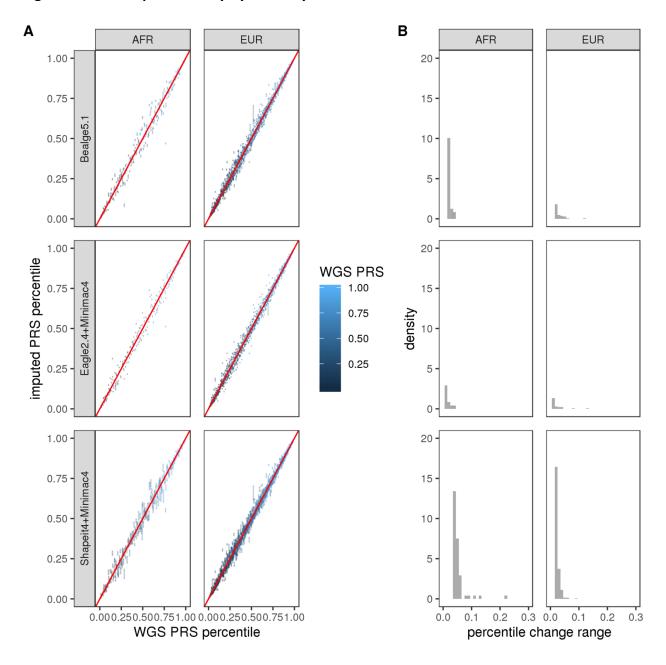
The variability in GPS_{BC} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S24. PRS-GWAS_{Afib} Reproducibility by Ancestry.



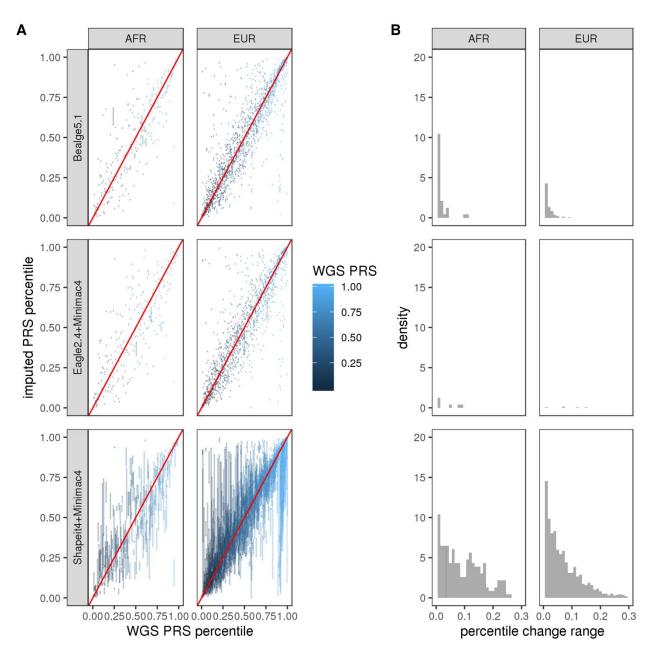
The variability in PRS-GWAS_{Afib} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S25. GPS_{Afib} Reproducibility by Ancestry.



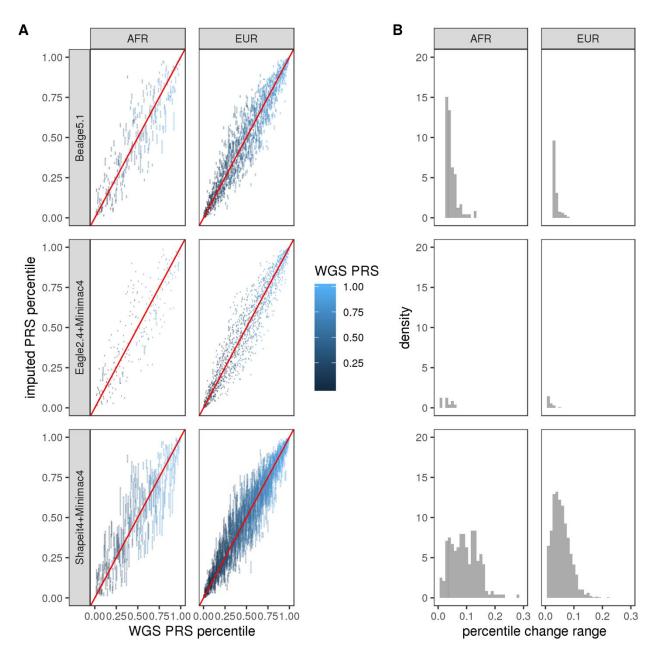
The variability in GPS_{Afib} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S26. PRS-GWAS_{AD} Reproducibility by Ancestry.



The variability in PRS-GWAS_{AD} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S27. PRS-GWAS_{Glaucoma} Reproducibility by Ancestry.



The variability in PRS-GWAS_{Glaucoma} percentile values as determined by three different imputation processes by 2 ancestries. **A.** Gold standard WGS-based PRS percentile (x-axis) vs six replicates of imputation derived PRS percentiles (y-axis). Point darkness depicts point density for overplotting. **B.** Histogram of the absolute score deviations relative to the WGS-based standard. Note, bin for no change is not shown. AFR: Afrian, EUR: European.

Fig S28. metaGRS_{CAD} Variability as a Function of PRS Bin.

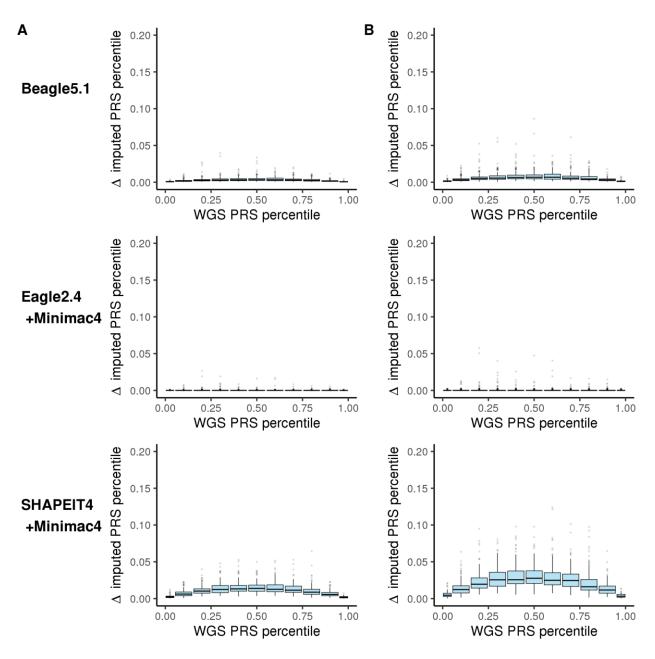


Fig S29. GPS_{CAD} Variability as a Function of PRS Bin.

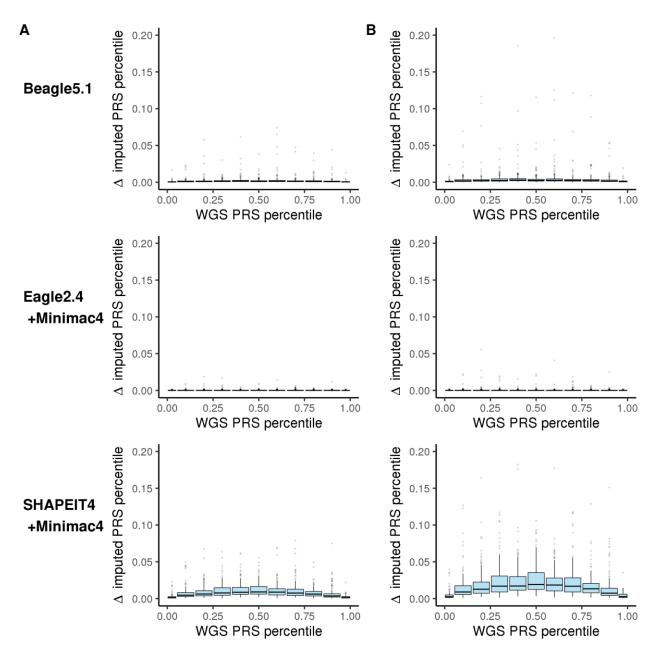


Fig S30. PRS-GWAS_{T2D} (547) Variability as a Function of PRS Bin.

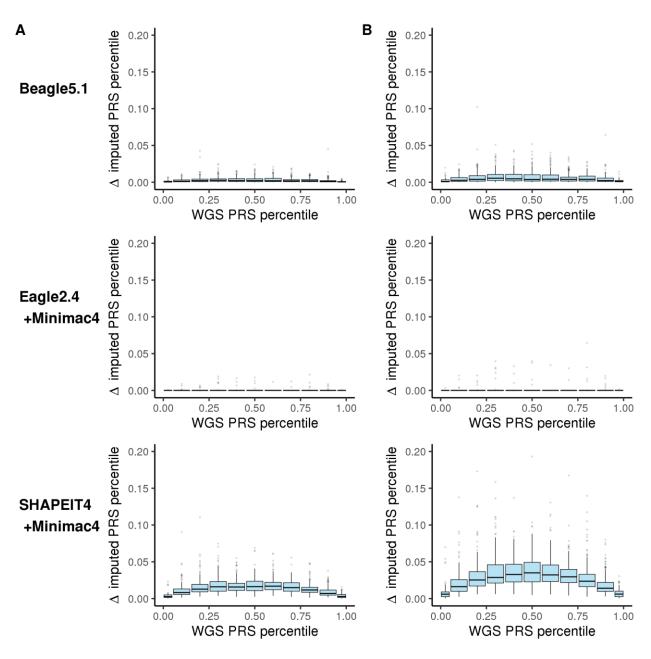


Fig S31. PRS-GWAS_{T2D} (397) Variability as a Function of PRS Bin.

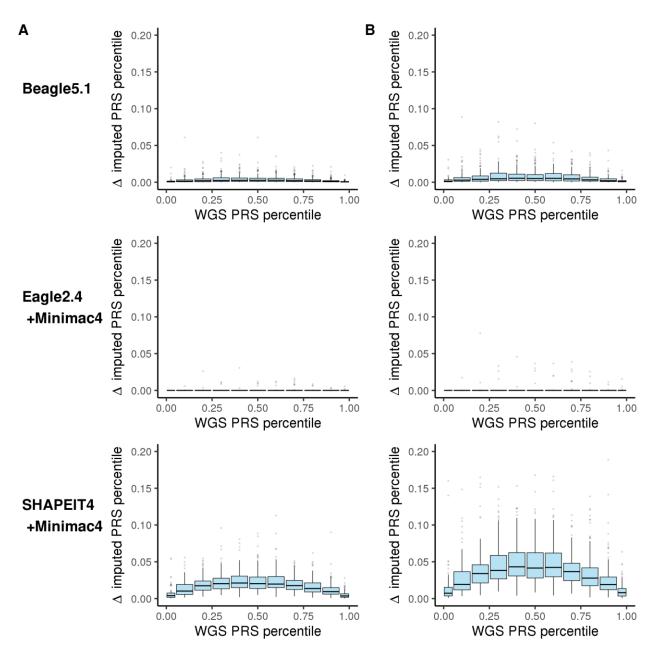


Fig S32. PRS-GWAS_{T2D} (170487) Variability as a Function of PRS Bin.

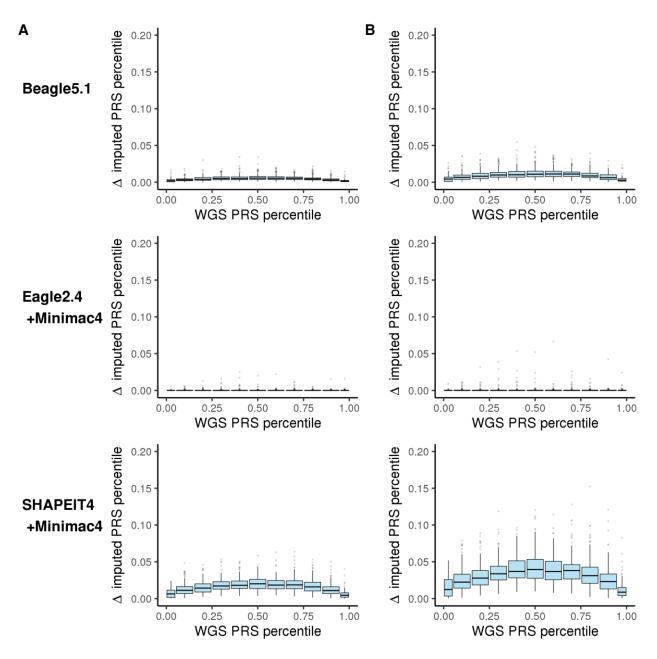


Fig S33. GPS_{T2D} Variability as a Function of PRS Bin.

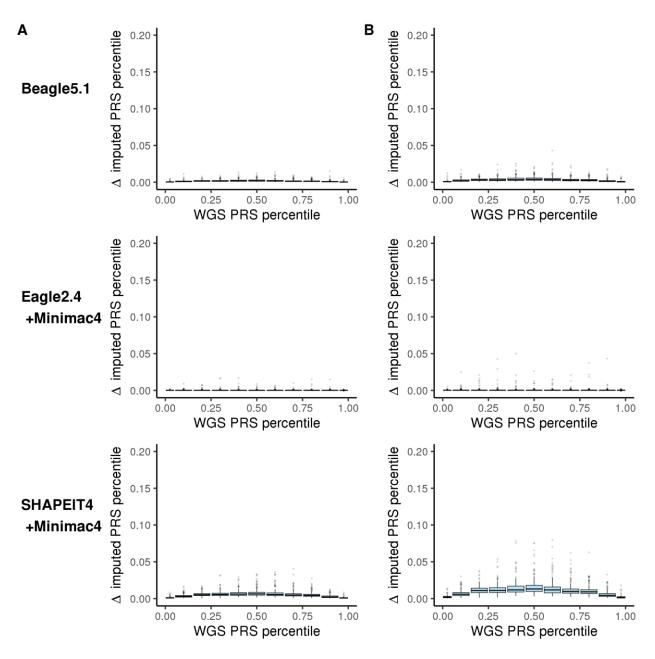


Fig S34. PRS-GWAS_{BC} (239) Variability as a Function of PRS Bin.

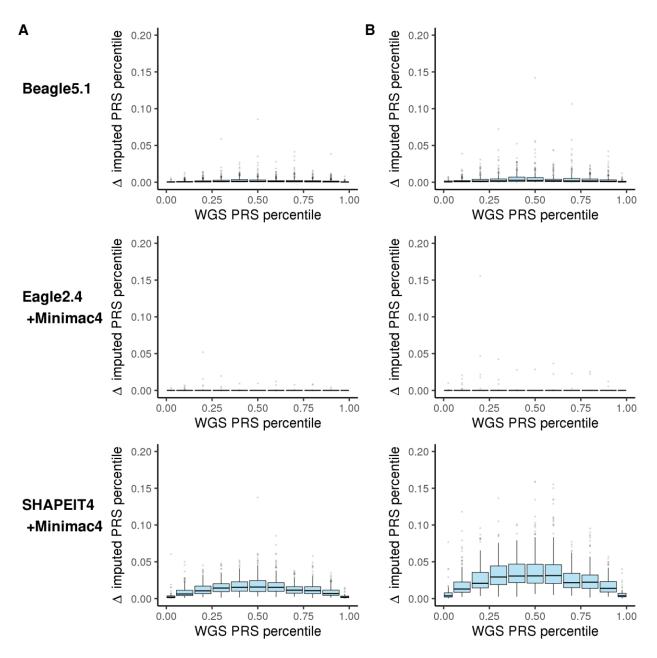


Fig S35. PRS-GWAS_{BC} (2935) Variability as a Function of PRS Bin.

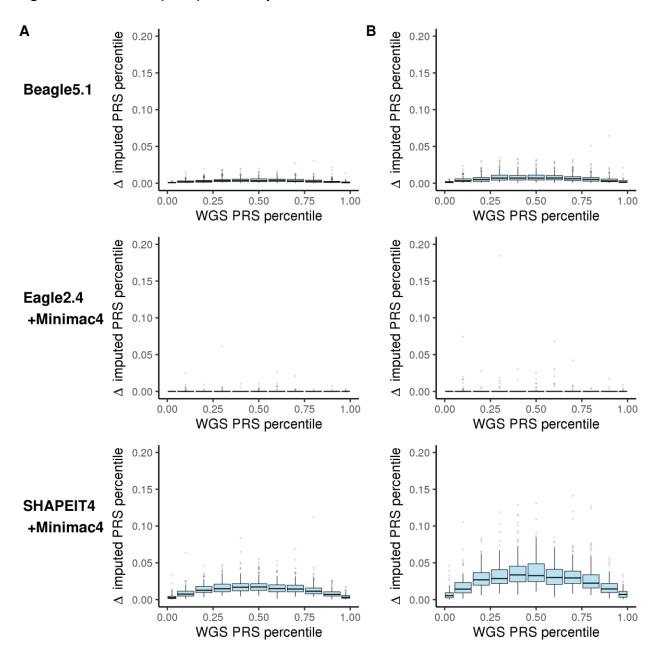


Fig S36. GPS_{BC} Variability as a Function of PRS Bin.

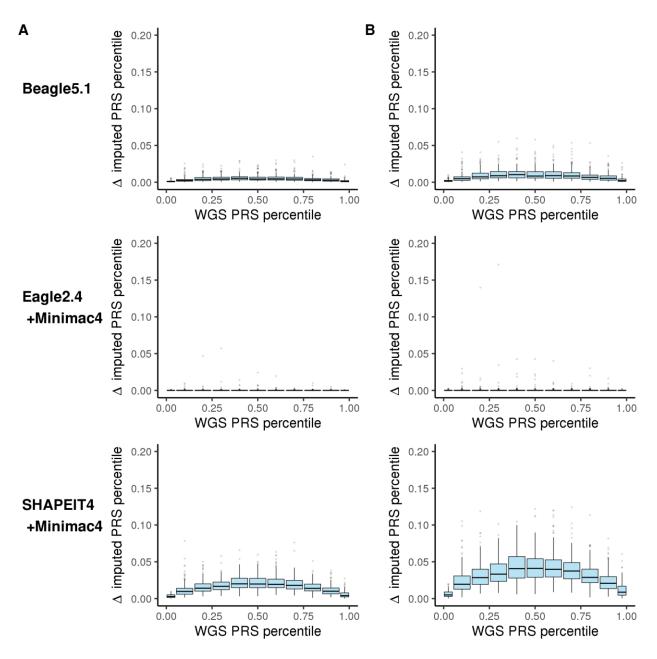


Fig S37. PRS-GWAS_{Afib} Variability as a Function of PRS Bin.

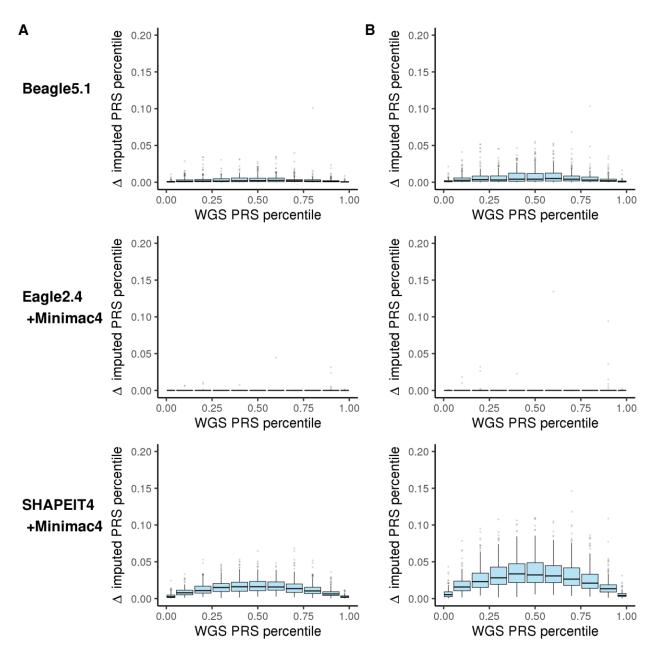


Fig S38. GPS_{Afib} Variability as a Function of PRS Bin.

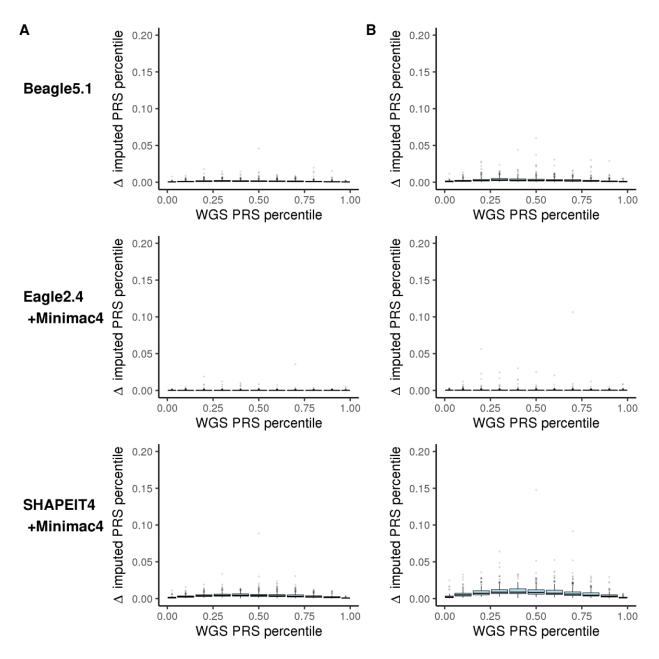


Fig S39. PRS-GWAS_{AD} Variability as a Function of PRS Bin.

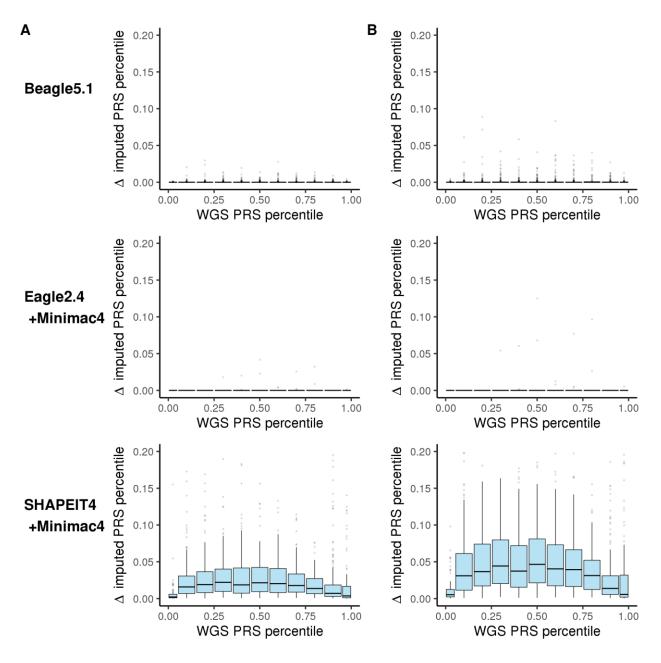


Fig S40. PRS-GWAS_{Glaucoma} Variability as a Function of PRS Bin.

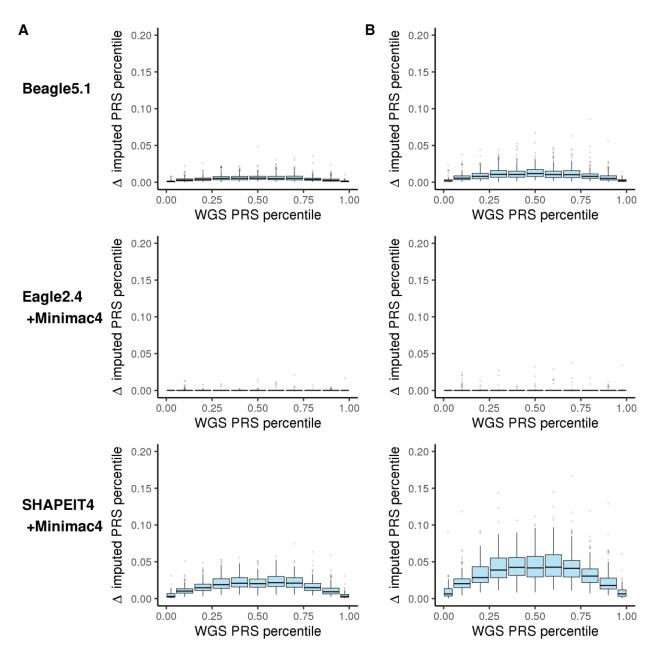


Fig S41. PRS_{CAD} Variability as a Function of PRS Bin by Ancestry.

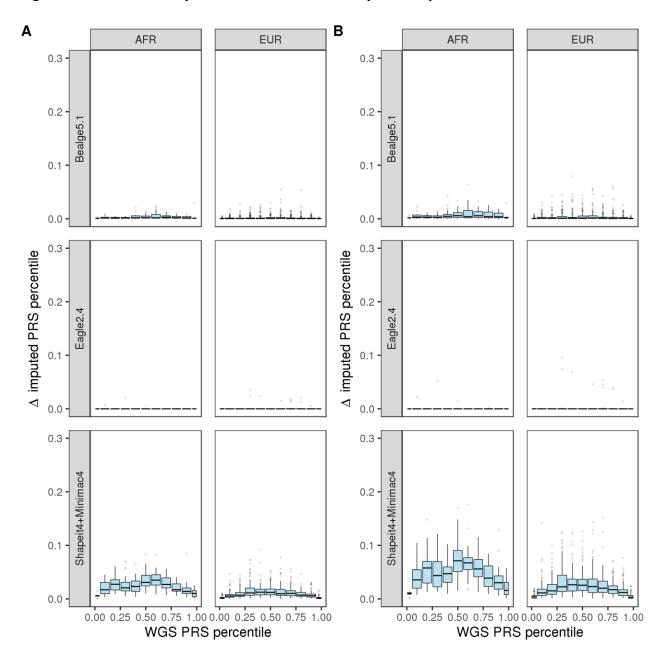


Fig S42. metaGRS_{CAD} Variability as a Function of PRS Bin by Ancestry.

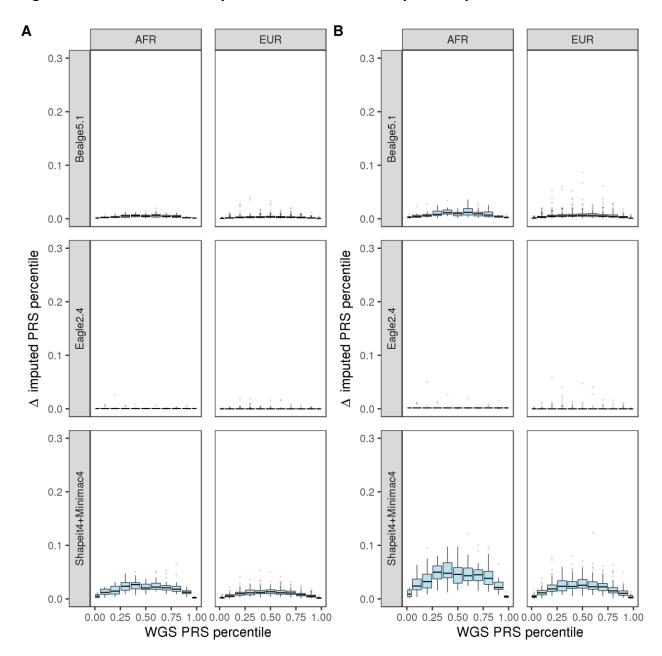


Fig S43. GPS_{CAD} Variability as a Function of PRS Bin by Ancestry.

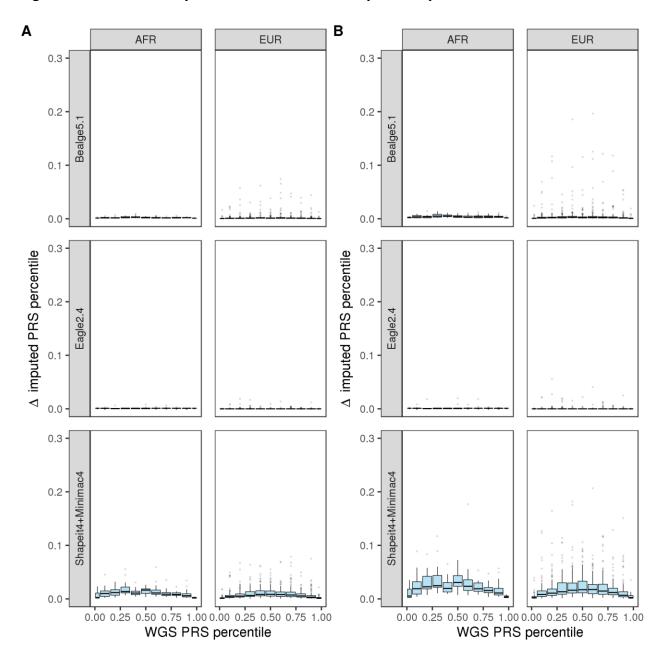


Fig S44. PRS-GWAS_{T2D} (547) Variability as a Function of PRS Bin by Ancestry.

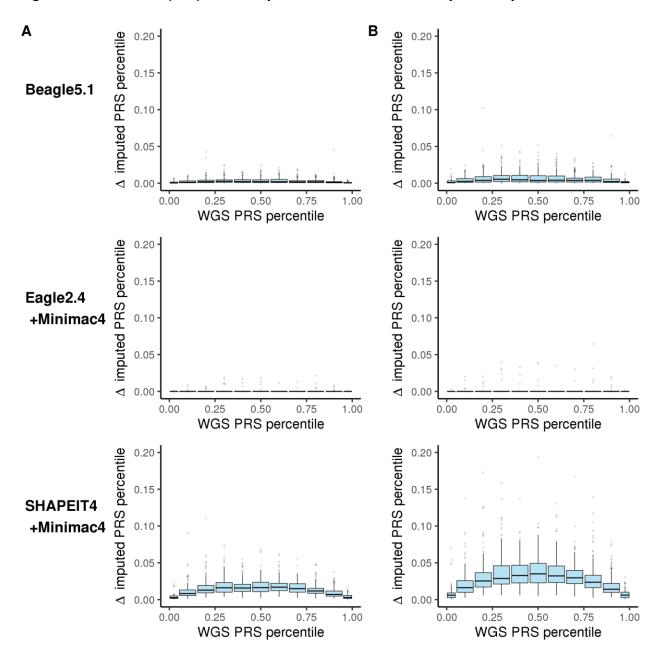


Fig S45. PRS-GWAS_{T2D} (397) Variability as a Function of PRS Bin by Ancestry.

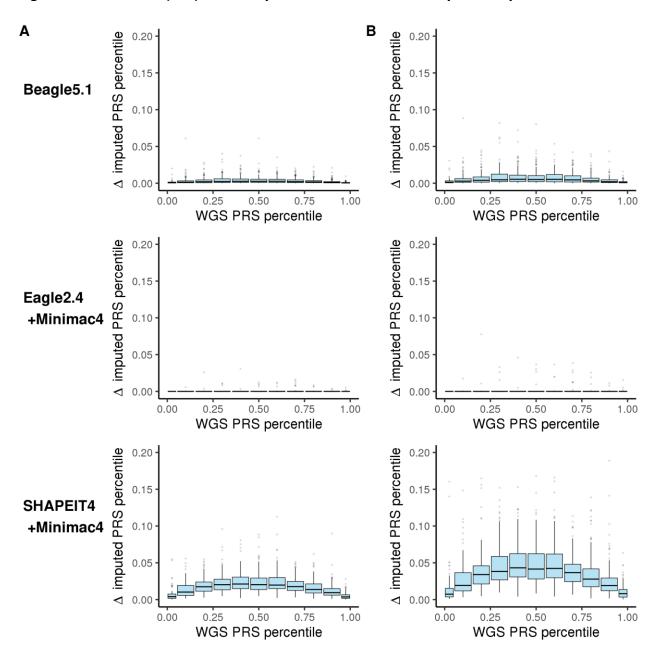


Fig S46. PRS-GWAS_{T2D} (170487) Variability as a Function of PRS Bin by Ancestry.

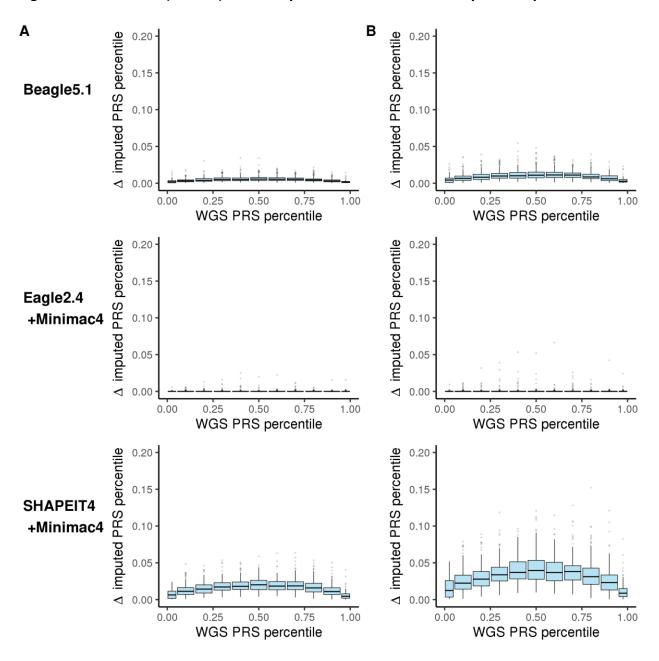


Fig S47. GPS_{T2D} Variability as a Function of PRS Bin by Ancestry.

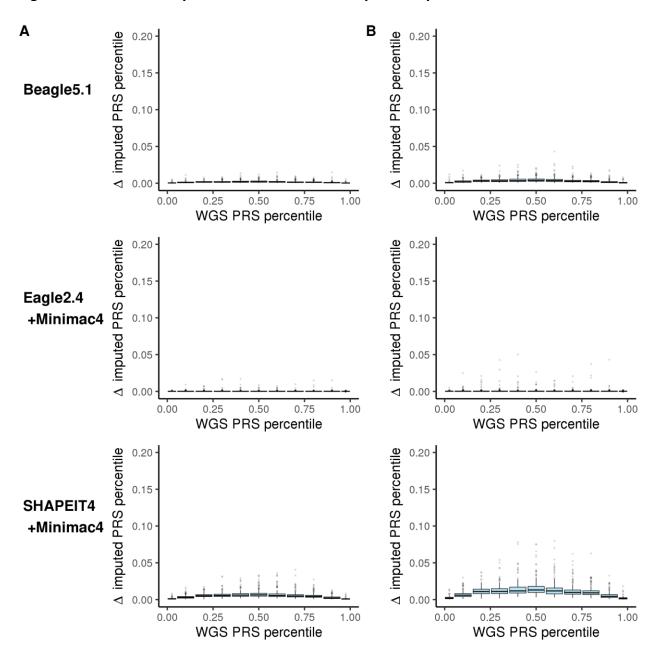


Fig S48. PRS-GWAS_{BC} (239) Variability as a Function of PRS Bin by Ancestry.

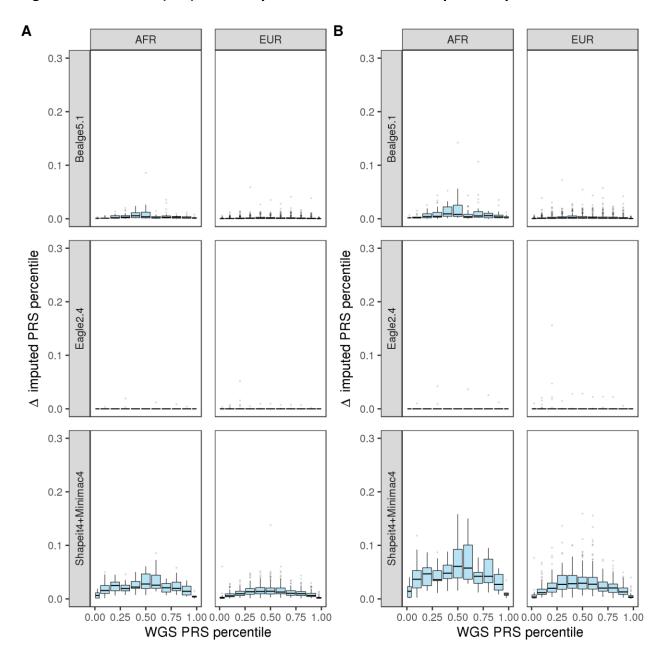


Fig S49. PRS-GWAS_{BC} (2935) Variability as a Function of PRS Bin by Ancestry.

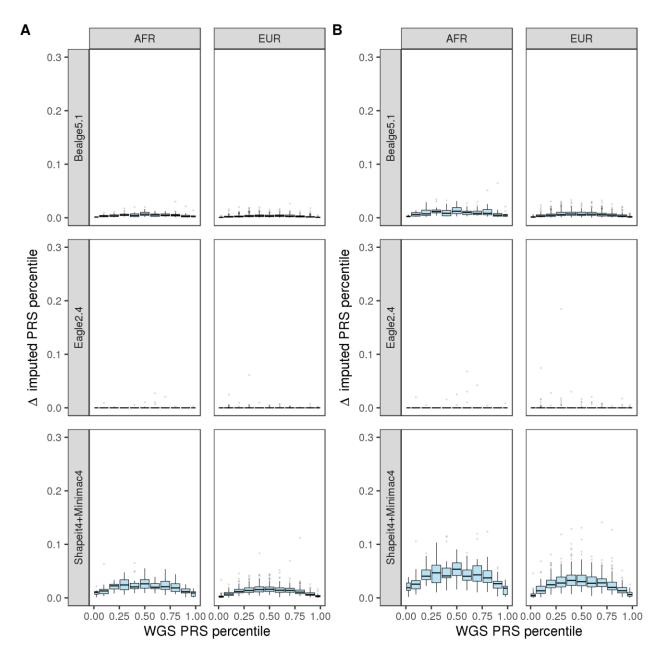


Fig S50. GPS_{BC} Variability as a Function of PRS Bin by Ancestry.

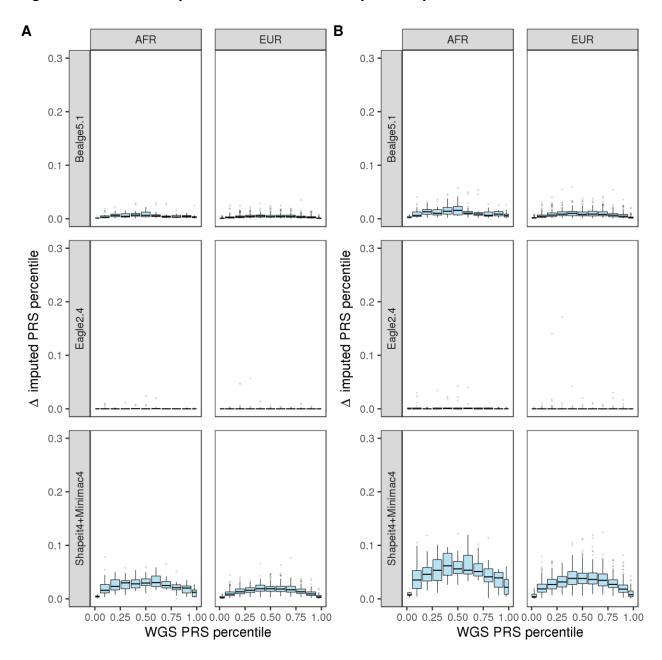


Fig S51. PRS-GWAS_{Afib} Variability as a Function of PRS Bin by Ancestry.

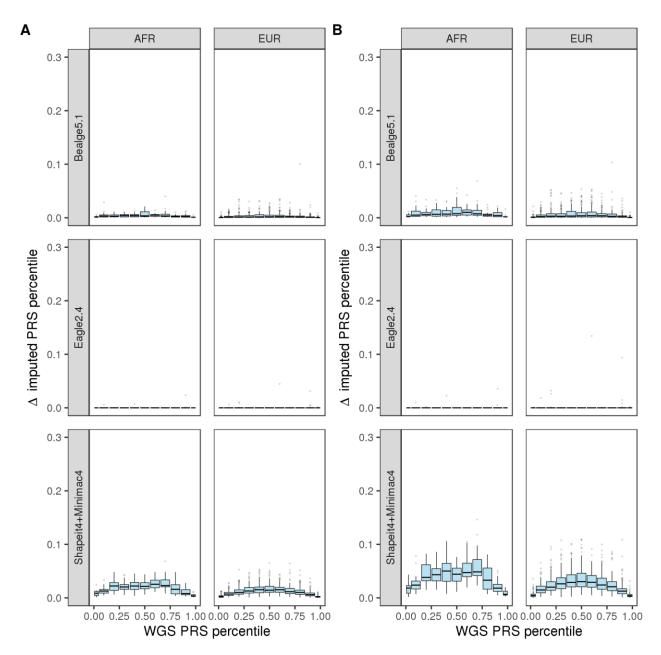


Fig S52. GPS_{Afib} Variability as a Function of PRS Bin by Ancestry.

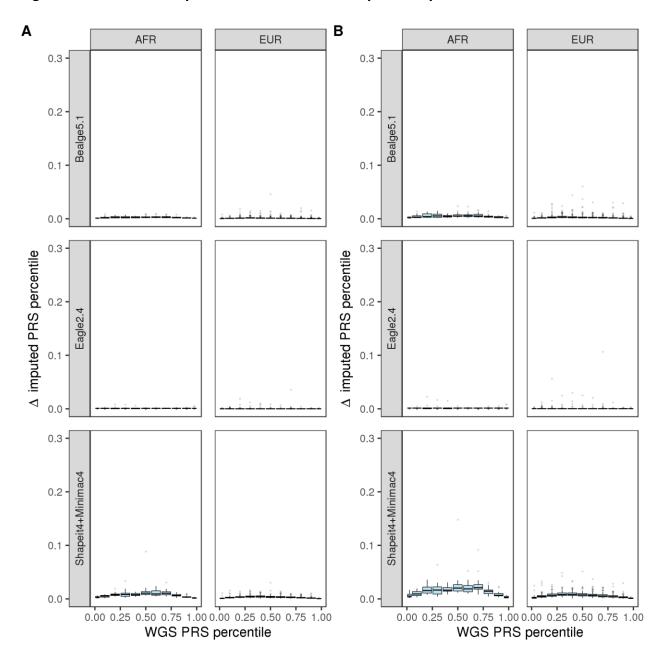


Fig S53. PRS-GWAS_{AD} Variability as a Function of PRS Bin by Ancestry.

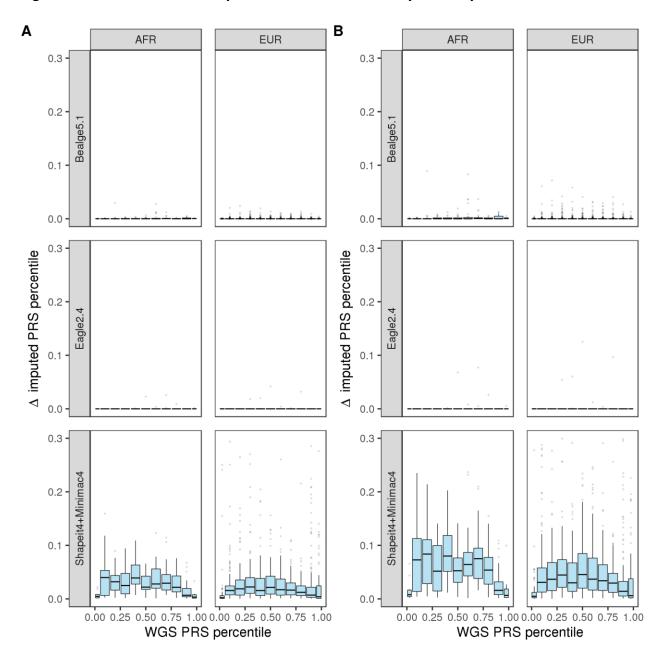
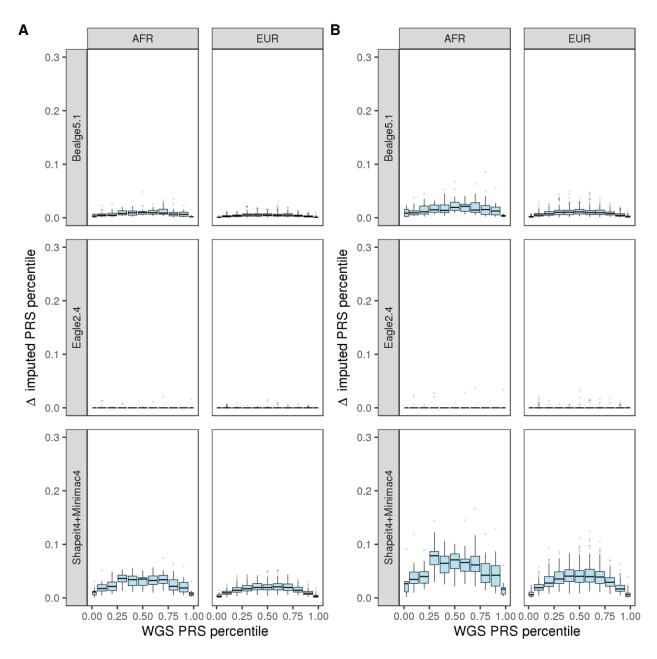


Fig S54. PRS-GWAS_{Glaucoma} Variability as a Function of PRS Bin by Ancestry.



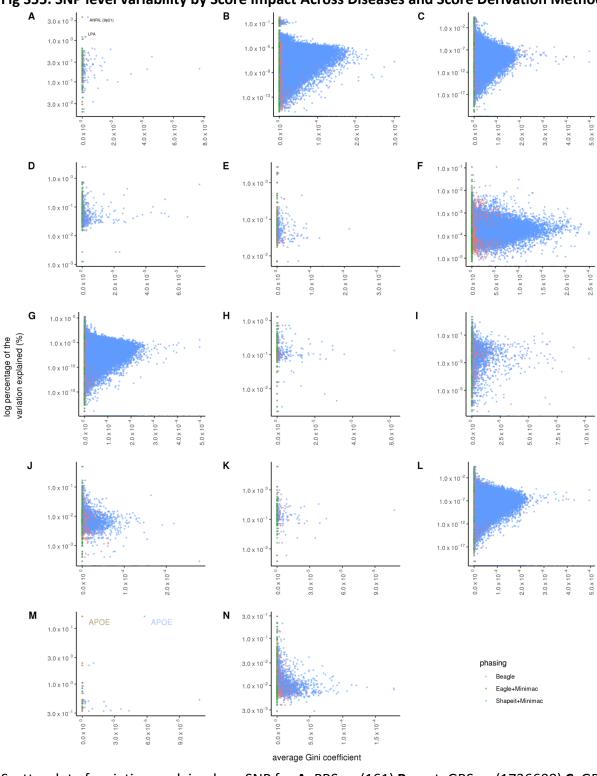


Fig S55. SNP level variability by Score Impact Across Diseases and Score Derivation Methods.

Scatterplot of variation explained per SNP for **A.** PRS_{CAD} (161) **B.** metaGRS_{CAD} (1736608) **C.** GPS_{CAD} (6238460) **D.** PRS-GWAS_{T2D} (547). **E.** PRS-GWAS_{T2D} (397). **F.** PRS-GWAS_{T2D} (170487). **G.** GPS_{T2D} (6482889). **H.** PRS-GWAS_{BC} (239). **I.** PRS-GWAS_{BC} (2941). **J.** GPS_{BC} (4457) **K.** PRS-GWAS_{Afib} (166). **L.** GPS_{Afib} (6302924). **M.** PRS-GWAS_{AD} (29). **N.** PRS-GWAS_{Glaucoma} (2657).